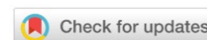
ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ
INFORMATION TECHNOLOGY, COMPUTER SCIENCE AND MANAGEMENT

УДК 004.9
https://doi.org/10.23947/2687-1653-2023-23-3-317-328

Научная статья

Метод трехмерного захвата движений человека на основе компьютерного зрения

А.Д. Обухов ✉, **Д.Л. Дедов**, **Е.О. Суркова**, **И.Л. Коробова**
Тамбовский государственный технический университет, г. Тамбов, Российская Федерация
✉ obuhov.art@gmail.com

Аннотация

Введение. Проведенный анализ существующих подходов к отслеживанию тела человека выявил наличие проблем при захвате движений в трехмерной системе координат. Отмечена перспективность систем захвата движений на основе компьютерного зрения. В существующих исследованиях по безмаркерным системам захвата движений рассматривается позиционирование только в двумерном пространстве. Поэтому целью исследования являлось повышение точности определения координат человеческого тела в трехмерных координатах за счет разработки метода захвата движения на основе компьютерного зрения и алгоритмов триангуляции.

Материалы и методы. Представлен метод захвата движений, включающий калибровку нескольких камер и формализацию процедур обнаружения человека в кадре с использованием сверточной нейронной сети. На основе полученных от нейронной сети скелетных точек осуществляется трехмерная реконструкция модели тела человека с использованием различных алгоритмов триангуляции.

Результаты исследования. Проведены экспериментальные исследования по сравнению четырех алгоритмов триангуляции: прямого линейного переноса, линейного метода наименьших квадратов, L2 триангуляции и полиномиального методов. Определен оптимальный алгоритм триангуляции (полиномиальный), обеспечивающий погрешность не более 2,5 пикселей или 1,67 сантиметров.

Обсуждение и заключение. Выявлены недостатки существующих систем захвата движения. Предложенный метод направлен на повышение точности захвата движений в трехмерных координатах с использованием компьютерного зрения. Полученные результаты интегрированы в программное обеспечение позиционирования тела человека в трехмерных координатах для удаленного мониторинга, использования в виртуальных тренажерах и системах захвата движений.

Ключевые слова: захват движений, виртуальная реальность, триангуляция, компьютерное зрения, машинное обучение

Благодарности: авторы выражают благодарность редакционной коллегии журнала и рецензенту за профессиональный анализ и рекомендации по корректировке текста статьи.

Финансирование. Исследование выполнено за счет гранта Российского научного фонда № 22–71–10057, https://rscf.ru/project/22-71-10057/

Для цитирования. Обухов А.Д., Дедов Д.Л., Суркова Е.О., Коробова И.Л. Метод трехмерного захвата движений человека на основе компьютерного зрения. *Advanced Engineering Research (Rostov-on-Don).* 2023;23(3):317–328. https://doi.org/10.23947/2687-1653-2023-23-3-317-328

Original article

3D Human Motion Capture Method Based on Computer Vision

Artem D. Obukhov ✉, **Denis L. Dedov**, **Ekaterina O. Surkova**, **Irina L. Korobova**
Tambov State Technical University, Tambov, Russian Federation
✉ obuhov.art@gmail.com

Abstract

Introduction. The analysis of approaches to tracking the human body identified problems when capturing movements in a three-dimensional coordinate system. The prospects of motion capture systems based on computer vision are noted. In existing studies on markerless motion capture systems, positioning is considered only in two-dimensional space. Therefore, the research objective is to increase the accuracy of determining the coordinates of the human body in three-dimensional coordinates through developing a motion capture method based on computer vision and triangulation algorithms.

Materials and Methods. A method of motion capture was presented, including calibration of several cameras and formalization of procedures for detecting a person in a frame using a convolutional neural network. Based on the skeletal points obtained from the neural network, a three-dimensional reconstruction of the human body model was carried out using various triangulation algorithms.

Results. Experimental studies have been carried out comparing four triangulation algorithms: direct linear transfer, linear least squares method, L2 triangulation, and polynomial methods. The optimal triangulation algorithm (polynomial) was determined, providing an error of no more than 2.5 pixels or 1.67 centimeters.

Discussion and Conclusion. The shortcomings of existing motion capture systems were revealed. The proposed method was aimed at improving the accuracy of motion capture in three-dimensional coordinates using computer vision. The results obtained were integrated into the human body positioning software in three-dimensional coordinates for use in virtual simulators, motion capture systems and remote monitoring.

Keywords: motion capture, virtual reality, triangulation, computer vision, machine learning

Acknowledgements: the authors would like to thank the editorial board of the journal and the reviewer for their professional analysis and recommendations for correcting the text of the article.

Funding information. The research was done on grant of the Russian Science Foundation No. 22–71–10057, https://rscf.ru/project/22-71-10057/

For citation. Obukhov AD, Dedov DL, Surkova EO, Korobova IL. 3D Human Motion Capture Method Based on Computer Vision. *Advanced Engineering Research (Rostov-on-Don).* 2023;23(3):317–328. https://doi.org/10.23947/2687-1653-2023-23-3-317-328

Введение. В области компьютерного зрения в настоящее время достигнут значительный прогресс. Разработаны технологии, позволяющие решать задачи обнаружения объектов, определения их состояния, геометрической оценки изображенного на кадре пространства и многие другие. Благодаря этому компьютерное зрение получило широкое распространение в различных сферах деятельности человека, начиная от здравоохранения и образования, заканчивая сферой развлечений. Достаточно перспективным направлением является применение технологий компьютерного зрения для трехмерной реконструкции и позиционирования различных объектов, в том числе и людей. Существует достаточно большое количество систем для определения абсолютного положения человека в пространстве, которые можно разделить по следующим категориям:

– системы, использующие инерциальные датчики и позволяющие определить величину их перемещения, а также изменение углов между ними, что предполагает использование гироскопов и акселерометров [1]. Известным представителем данной категории является Noitom Mocap Perception Neuron [2], включающий до 32 инерциальных датчиков;

– лазерные системы позиционного слежения, основанные на использовании базовых станций, установленных на противоположных сторонах помещения и излучающих инфракрасные лучи, которые позволяют точно определять положение и ориентацию датчиков в пространстве. Примером таких систем являются комплекты виртуальной реальности от компании HTC [3], обладающие погрешность до 0,1 мм;

– системы, использующие магнитные датчики [4], основанные на применении магнитного поля для захвата движения человека, которые предполагают наличие носимых датчиков на теле пользователя. К данной категории относится Polhemus Liberty — портативная электромагнитная система отслеживания движений, считающаяся одной из самых быстрых (частота дискретизации 240 Гц);

– оптические системы на основе маркеров — определяют положение объектов по маркерам с использованием набора камер. Примером является Vicon, обладающий достаточно низкой погрешностью: средние абсолютные ошибки отслеживания маркера составляют 0,15 мм при статических испытаниях и 0,2 мм (с соответствующими угловыми ошибками 0,3°) — при динамических испытаниях [5];

– безмаркерные оптические системы, основанные на использовании компьютерного зрения и машинного обучения. Примерами таких технологий являются OpenPose, MediaPipe, Movenet. С их помощью движения человека могут отслеживаться с точностью до 30 мм [6].

Анализируя перечисленные категории систем захвата движений, можно сделать вывод, что большинство решений, используемых для распознавания человеческих действий и движений предполагает наличие различных носимых устройств, таких как сенсоры или перчатки. Основная масса этих устройств являются громоздкими из-за большого количества датчиков и необходимости проводного подключения. Некоторые такие системы обладают высокой точностью, но при этом не могут быть использованы из-за размеров или наличия электромагнитных помех [7]. Инерциальные системы имеют ряд проблем, связанных с накоплением ошибок, что ограничивает их использование только относительным позиционированием в пространстве.

Поэтому большой популярностью пользуются оптические системы для распознавания и отслеживания действий пользователя. Для получения информации о действиях и положении пользователя используются кадры, полученные с камеры. Среди оптических систем стоит отметить те, которые используют маркеры (пользователь может быть одет в специальную одежду или на нем закреплены определенные метки), что затрудняет их использование в реальных условиях и больше применимо для специально подготовленных помещений (например, киностудий).

Системы, не использующие какие-либо маркеры, позволяют пользователям более свободно взаимодействовать с окружающей средой и больше подходят для применения в реальных условиях. К существенным недостаткам систем этого направления можно отнести относительно невысокую точность, ненадёжность и низкую производительность. Во многом это может быть обусловлено недостатками алгоритмов компьютерного зрения, используемых для распознавания человека в кадре, а также следующими причинами: изменчивостью внешнего вида человека и условий освещения, частичными окклюзиями из-за наслоения объектов в сцене, сложностью человеческой скелетной структуры.

В основе работы безмаркерных систем захвата движения, как правило, используется алгоритм оценки позы человека. Подходы к решению задачи оценки позы человека можно разделить на нисходящие и восходящие. В нисходящих подходах сначала происходит обнаружение людей в кадре, затем оценка позы каждого найденного человека. Алгоритмы, которые относятся к восходящим подходам, на первом этапе выполняют поиск частей тела в кадре, затем производят их группировку в позы. Как правило, для этой задачи применяются сверточные нейронные сети, такие как YOLO (You Look Only Once) [8], SSD (Single Shot Detection) [9], R–CNN (Region CNN) [10] и другие. Они позволяют распознавать множество различных объектов, включая человека или отдельные части тела с высокой точностью. Однако одним из недостатков решений, перечисленных выше, можно назвать их низкую производительность и медленную работу. Для решения этой проблемы существуют специальные фреймворки (MoveNet [11], MediaPipe [12], OpenPose [13]), также использующие нейронные сети, оптимизированные для работы в режиме реального времени.

Необходимо отметить, что вышеуказанные алгоритмы, технологии и подходы безмаркерных систем захвата движения позволяют осуществить позиционирование в двумерном пространстве, что затрудняет как определение расстояния до объектов и их размеров, так и отслеживание сложных движений, когда, например, руки пользователя скрыты его телом. Существующие решения в области стереокамер могут быть эффективны, но не отличаются большой точностью при значительном удалении объекта от камеры, что происходит при отслеживании всего тела человека. Кроме того, они не решают проблемы окклюзий. Таким образом, актуальным направлением исследований является разработка метода захвата движений с использованием нескольких камер и технологий компьютерного зрения. При реализации многокамерных систем захвата движений неизбежно возникает проблема совмещения объектов с нескольких изображений, т.е. необходимость выполнения триангуляции. Среди методов триангуляции можно выделить линейные и итеративные линейные алгоритмы.

Линейная триангуляция наиболее распространённый подход к выполнению реконструкции объектов в трёхмерном пространстве, включающая такие методы как линейно-собственный метод, линейный метод наименьших квадратов, прямое линейное преобразование, отличающиеся разной степенью устойчивости к шуму [14].

Итеративные линейные методы представляют собой более надёжную версию линейной триангуляции. Обычные линейные методы могут быть менее точны при решении задач триангуляции совокупности точек, так как при решении систем минимизируемая ошибка не имеет геометрического смысла (не учитывает форму

скелета и правила соединения точек). Основная идея итерационных линейных методов заключается в адаптивном изменении весов линейных уравнений таким образом, что взвешенные уравнения соответствуют ошибкам. К итеративным линейным методам можно отнести L2 и L∞ триангуляцию [15].

Таким образом, в рамках данного исследования ставится следующая задача: необходимо разработать метод захвата движений человека, позволяющий позиционировать тело пользователя в трехмерных координатах с минимальной погрешностью и с использованием технологий компьютерного зрения. Предлагаемый метод может использоваться как замена существующих систем захватам движений, так и входить в состав других алгоритмов, например, для последующей классификации состояния человека. Целью данной работы является повышение точности определения поз и координат человеческого тела в трехмерных координатах за счет разработки методов захвата движения на основе компьютерного зрения. Для достижения поставленной цели необходимо формализовать основные этапы процесса захвата точек тела человека с нескольких камер, интегрировать алгоритмы триангуляции, выбрав среди них оптимальный с точки зрения точности, осуществить программную реализацию предложенного метода.

Материалы и методы. Решение задачи трехмерного позиционирования человека в пространстве включают следующие основные этапы:

– предварительная калибровка набора камер;

– реализация процедур обнаружения человека в кадре и расчет скелетных точек;

– расчет трехмерной реконструкции модели тела человека.

Рассмотрим их более подробно.

Процесс калибровки включает в себя выполнение системой камер нескольких снимков калибровочного шаблона, на котором можно легко выделить ключевые точки с известными для них относительными положениями в пространстве. После осуществляется расчет внутренних и внешних параметров для каждой камеры. Внутренние параметры постоянны для конкретной камеры, внешние — зависят от расположения камер относительно друг друга [16]. Поэтому данный шаг необходимо выполнять перед первым использованием системы камер в заданном расположении.

Для вычисления значений координат точки в трёхмерном пространстве необходимо знать координаты её проекций на изображениях и проективные матрицы камер [17]. Проективная матрица P некоторой камеры может быть представлена в виде комбинации матриц A (содержащей внутренние параметры камеры) и R (вращения), а также вектора перемещения T, которые описывают изменение координат из мировой системы координат в систему координат относительно камеры:

$$P = A[R \mid T] = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}, \quad (1)$$

где (x, y) — координаты проекции трёхмерной точки на изображении в пикселях; (c_x, c_y) — координаты центральной точки камеры; (f_x, f_y) — фокусное расстояние в пикселях.

На втором этапе необходимо получить непосредственно ключевые (скелетные) точки тела человека на каждой их камер. Для извлечения скелетных точек тела из кадра возможно применение различных технологий машинного обучения, например, MoveNet, MediaPipe, OpenPose и других [18]. В рамках данного исследования предлагается использовать высокоэффективный и производительный модуль Pose из библиотеки MediaPipe. MediaPipe Pose использует машинное обучение для высокоточного отслеживания позы тела человека, определения трехмерных ориентиров и маски сегментации фона на всем теле из видеокадров RGB. Данный подход позволяет отслеживать до 33 точек и обеспечивает работу в реальном времени на большинстве современных устройств.

Таким образом, в рамках второго этапа формируется набор из 33 точек для каждой *i*-ой камеры:

$$\left\{ x_{ij} = \left\langle u_{ij}, v_{ij} \right\rangle \mid j \in \{1, 2, ..., 33\}, i \in \{1, 2, ..., K\} \right\}, \quad (2)$$

где u_{ij} — координата *j*-ой точки по оси X на *i*-ом изображении; v_{ij} — координата *j*-ой точки по оси Y на *i*-ом изображении; K — общее количество камер и изображений.

На третьем этапе выполняется расчёт положений ключевых скелетных точек в трёхмерном пространстве. Для получения данных о положении скелетных точек человека в пространстве выполняется триангуляция — нахождение координат трёхмерной точки по координатам её проекций. Триангуляция — одна из важнейших задач в компьютерном зрении, её решение является решающим этапом в 3D-реконструкции и влияет на точность всего результата [19].

В основе трехмерной реконструкции точек объекта по значениям положений проекций точек на изображениях со всех камер лежит эпиполярная геометрия. Основная её идея заключается в том, что 3D-точки в сцене проецируются на линии в плоскости изображения каждой камеры — эпиполярные линии. Эти линии соответствуют пересечению плоскости изображения с плоскостью, проходящей через центры камер и 3D-точку. Данная идея обеспечивает условие для поиска пар соответствующих точек на двух изображениях: если известно, что точка x на плоскости первого изображения соответствует точке x' на плоскости другого изображения, то ее проекция должна лежать на соответствующей эпиполярной линии. Согласно этому условию, для всех соответствующих пар точек $x \leftrightarrow x'$ будет справедливо следующее отношение:

$$x' F x = 0, \tag{3}$$

где F — фундаментальная матрица, имеющая размер 3×3 и ранг равный 2.

Для некоторой точки X, заданной в трёхмерном пространстве, справедлива следующая формула проецирования, выражающаяся в однородных координатах:

$$x_i = P_i X, \tag{4}$$

где $x_i = w(u_i, v_i, 1)^T$ — однородные координаты некоторой точки на плоскости *i*-го изображения (полученной с *i*-ой камеры в ходе второго этапа), включающие положение на изображении u_i (по оси X) и v_i (по оси Y); w — коэффициент масштабирования; P_i — проекционная матрица *i*-й камеры, полученная на первом этапе.

Для упрощения вычислений проекционная матрица камеры зачастую представляется в следующем виде:

$$P_i = \begin{bmatrix} p_i^{1T} \\ p_i^{2T} \\ p_i^{3T} \end{bmatrix} \left(P_i \in \mathbb{R}^{3\times4} \right), \tag{5}$$

где p_i^{jT} — *j*-ая строка матрицы P_i.

Следовательно, уравнение (4) может быть представлено следующим образом:

$$\begin{aligned} wu_i &= p_i^{1T} X, \\ wv_i &= p_i^{2T} X, \\ w &= p_i^{3T} X. \end{aligned} \tag{6}$$

Так как w — коэффициент масштабирования, получаем следующую систему уравнений:

$$\begin{aligned} u_i p_i^{3T} X - p_i^{1T} X &= 0, \\ u_i p_i^{3T} X - p_i^{2T} X &= 0. \end{aligned} \tag{7}$$

Так как X является однородным представлением координат в трёхмерном пространстве, то для их вычисления необходимо получить x_i и P_i как минимум для двух камер. Для решения системы уравнений (7) рассмотрено 4 алгоритма [14]:

- прямой линейный перенос (DLT);
- линейный метод наименьших квадратов;
- L2 триангуляция;
- оптимальный (полиномиальный) метод.

DLT относится к линейным алгоритмам триангуляции, главным достоинством которого является простота его реализации. Так, например, в библиотеке компьютерного зрения OpenCV существует готовая реализация данного алгоритма в методе triangulatePoints.

Линейный метод наименьших квадратов также относится к линейным и состоит в том, что система однородных уравнений (7) сводится к системе, состоящей из неоднородных уравнений, для решения которой используется метод наименьших квадратов.

L2 триангуляция является итеративным методом трехмерной реконструкции, решение которой сводится к минимизации ошибки перепроецирования:

$$\sum_i d(x_i, \widehat{x_i}) \to \min, \tag{8}$$

где x_i — координата проекции оцениваемой точки на изображении; $\hat{x}_i$ — координата проекции, вычисленная по формуле (4) для уже рассчитанной пространственной точки; $d(\bullet)$ — расстояние между двумя точками.

Алгоритм оптимальной (полиномиальной) триангуляции относят к неитеративным подходам. Для его решения требуется полином шестого порядка. Критерий минимизации для выполнения трёхмерной реконструкции в данном методе можно определить следующим образом:

$$\sum_i d(x_i, \lambda_i) \to \min, \qquad (9)$$

где λ_i — эпиполярная линия, соответствующая точке x_i.

При использовании системы из двух камер для минимизации ошибки (9) необходимо выполнить следующую последовательность действий:

– параметризировать пучок эпиполярных линий на первом изображении с помощью параметра t. Таким образом эпиполярная линия на первом изображении может быть выражена как $\lambda_0(t)$;

– используя фундаментальную матрицу F, вычислить соответствующую эпиполярную линию $\lambda_1(t)$ на втором изображении;

– выразить функцию расстояния (9) как функцию от t;

– выполнить поиск значения t, при котором (9) стремится к минимуму.

Используя методы элементарного исчисления, можно свести решение задачи минимизации к нахождению корней полинома шестого порядка. Вычисление предполагаемой пространственной точки осуществляется с помощью метода прямого линейного переноса (DLT) [17].

Подводя итог третьему этапу, получим, что после успешного решения системы (7) и получения мировых координат ключевых точек целевого объекта (тела человека) формируется следующее множество точек H:

$$H = \left\{ X_j \mid \forall i \left(x_{ij} = P_i X_j \right) \right\}, \qquad (10)$$

где X_j — мировые координаты скелетной точки тела человека, полученные после решения задачи триангуляции, выраженные в сантиметрах.

Таким образом, задача оптимизации в данном исследовании при использовании двух камер сводится к нахождению такого метода триангуляции $MT : \{x_{ij}\} \to H$, при котором ошибка перепроецирования стремится к минимуму:

$$R = \frac{\sum_{i=1}^{2} \sum_{j=1}^{K} d(x_{ij}, \widehat{x_{ij}})}{K} \to \min. \qquad (11)$$

Результаты исследования. Решение задачи оптимизации (11) осуществляется при выполнении триангуляции двумерных точек объекта, полученных с изображений нескольких камер, в рамках данного исследования — с двух камер с использованием различных алгоритмов, перечисленных в предыдущем разделе.

Перечисленные методы триангуляции были реализованы с использованием библиотек OpenCV и NumPy. Для сравнения алгоритмы были интегрированы в программное обеспечение, реализующее метод трёхмерного захвата движений. Пример работы метода для реконструкции всего скелета человека представлен на рис. 1.

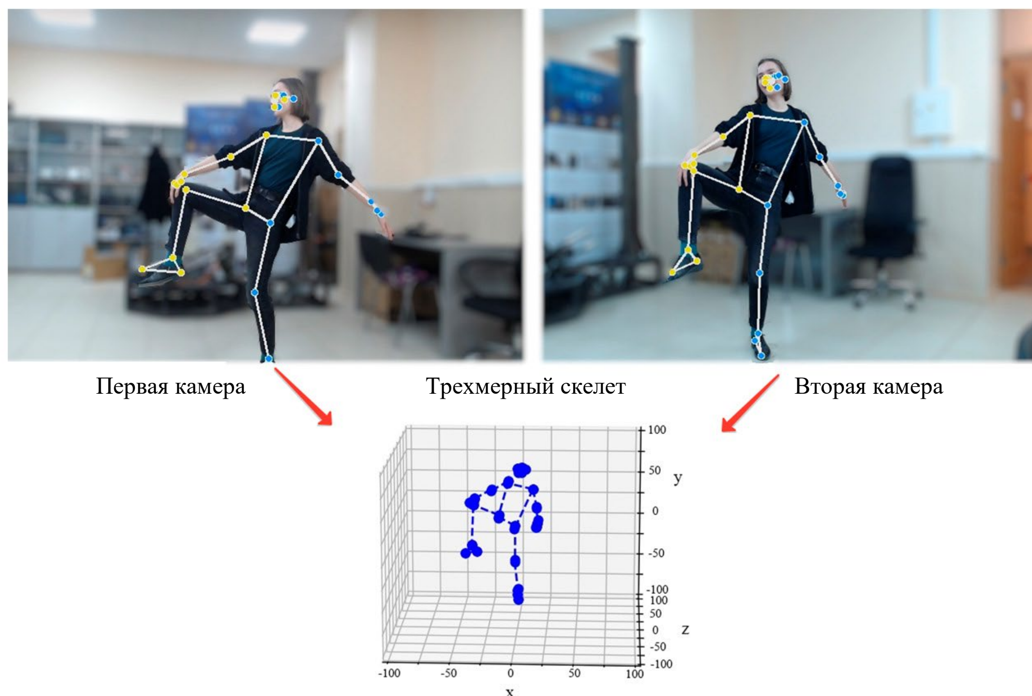


Рис. 1. Пример работы метода, включающий распознавание человека на двух камерах и построение трёхмерного скелета

Затем было выполнено сравнение данных алгоритмов по значению функции ошибки перепроецирования (11) для всех точек скелета с двух изображений. Проведено сравнение выбранных методов триангуляции по величине ошибки, а также по времени получения решения (вычислительной сложности) для всей совокупности точек скелета. Сводные сравнительные диаграммы представлены на рис. 2.

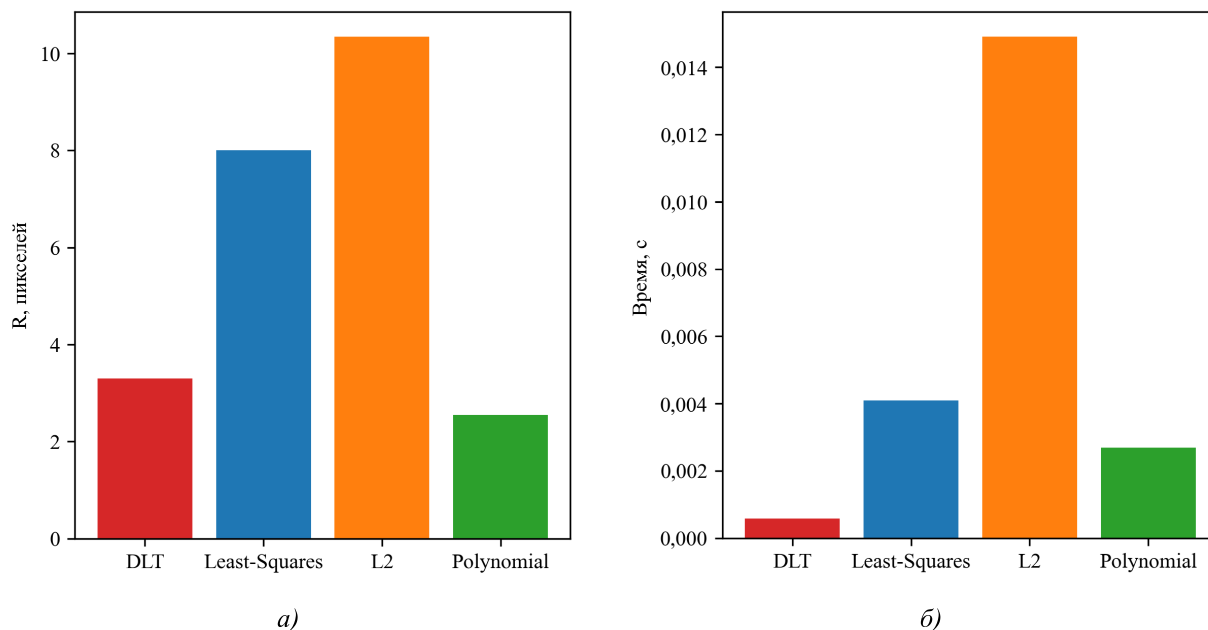


Рис. 2. Сравнение методов триангуляции по метрикам: *а* — по ошибке перепроецирования; *б* — по времени вычислений

Для выбранных методов триангуляции также проведен ряд экспериментальных тестов, в ходе которых для каждого подхода замерялись рассчитанные длины конечностей пользователя и абсолютное отклонение полученных значений от реальных. Сравнение представлено в таблице 1.

Таблица 1

Сравнение точности определения размеров конечностей в процессе триангуляции

Сегмент тела	DLT	Least-Squares	L2	Polynomial	Реальное значение
Предплечье	25,2 ± 1,6	30,8 ± 0,2	26,6 ± 0,5	24,3 ± 0,4	26
Голень	42,2 ± 2,0	65,3 ± 1,1	44,6 ± 0,7	38,7 ± 1,8	41
Бедро	45,7 ± 2,7	59,5 ± 0,49	48,7 ± 1,3	44,1 ± 0,6	45
Среднее отклонение	2,43	14,58	2,26	1,67	0
Представлены средние значения (в сантиметрах) после в выборке из 10 измерений ± среднее отклонение в выборке					

Разработанное программное обеспечение включает следующие модули:
– для работы с устройствами ввода (камерами);
– для выполнения калибровки и получения основных параметров камер;
– для синхронизации нескольких камер;
– для распознавания объектов (тело и руки пользователя);
– для анализа расположения найденных скелетных точек;
– для построения визуализации в режиме реального времени.

При реализации программного обеспечения использовался язык программирования Python, библиотеки OpenCV и Matplotlib. Функционирование системы осуществляется в нескольких потоках: один отвечает за получение данных с камер, второй — за визуализацию, третий — за отправку полученных мировых координат тела человека во внешние системы или модули. Использование унифицированного протокола с пакетом данных в формате JSON позволяет интегрировать программное обеспечение в сторонние системы (например, игровые среды разработки Unity, Unreal Engine и т.д.) [20, 21].

Обсуждение и заключение. Проанализируем результаты сравнения алгоритмов триангуляции по выбранным метрикам, представленные на рис. 2 и в таблице 1.

В ходе сравнения обнаружено, что оптимальным алгоритмом для трехмерной реконструкции является полиномиальный метод. Значение ошибки составляет порядка 2,55 пикселей. В реальных испытаниях при определении роста человека ошибка составила не более 3 %, с учетом того, что MediaPipe Pose не фиксирует верхнюю точку головы и она рассчитывается примерно на основе положения глаз. При измерении конечностей погрешность составила от 0,9 см до 2,3 см, средняя — 1,67 (таблица 1). Таким образом, реальные испытания подтверждают правильность выбора полиномиального метода.

Далее сравним полученные результаты с существующими исследованиями, например, описанными в работе [22]. Авторы также используют обученные сети (OpenPose) для реализации безмаркерной системы распознавания человека, процедуру калибровки камер, извлечение скелетных точек, но располагают камеры рядом друг с другом для имитации стереозрения. Это ключевое отличие позволяет распознавать в рамках данного исследования позы человека, когда одни части тела перекрывают другие. Кроме того, использование MediaPipe Pose позволяет отслеживать 33 скелетные точки, а не 18, как в методе на основе OpenPose. Полученные значения ошибок в целом соответствуют существующим исследованиям (наилучший результат в работе [22] составляет 2 см), что позволяет сделать вывод о возможности практического использования предложенного подхода. Прочие безмаркерные системы, например, на основе Kinect [23], также показывают сравнимые результаты по величине погрешности измерений (2–5 см). Таким образом, полученное решение в целом соответствует по точности существующим разработкам.

Сравнение времени вычисления набора точек, представленное на рис. 2 справа, показывает, что алгоритм DLT обеспечивает наибольшую производительность, однако, все алгоритмы показывают приемлемые результаты (для обеспечения быстродействия в 30 и даже 60 кадров в секунду). Поэтому данная метрика не является определяющей.

Разработанное программное обеспечение может использоваться в различных предметных областях прежде всего, как замена систем захвата движений, основанных на инерциальных датчиках. Преимуществами предлагаемого решения являются низкие экономические затраты на реализацию и доступность (переход от узкоспециализированных костюмов motion capture к распространенным средствам на основе камер), возможность параллельного захвата моделей тел нескольких пользователей [24].

Научная новизна исследования состоит в комплексном подходе к формализации процесса трехмерного позиционирования человека с использованием технологий компьютерного зрения, включая предварительную калибровку набора из нескольких камер, формализацию процедур обнаружения человека в кадре с использованием произвольной нейронной сети для получения скелетных точек, а также расчет трехмерной реконструкции модели тела человека с использованием различных алгоритмов триангуляции. Исследование включает все необходимые расчетные формулы и подробно изложенные этапы, позволяющие достичь поставленной цели — повышения точности определения поз и координат человеческого тела в трехмерных координатах с использованием технологий компьютерного зрения. Представленные теоретические результаты достаточно универсальны и могут быть использованы для практической реализации систем захвата движений, основанных на различных моделях нейронных сетей, а не только MediaPipe Pose.

Список литературы

1. Lind C.M., Abtahi F., Forsman M. Wearable Motion Capture Devices for the Prevention of Work-Related Musculoskeletal Disorders in Ergonomics – An Overview of Current Applications, Challenges, and Future Opportunities. *Sensors.* 2023;23(9):4259. https://doi.org/10.3390/s23094259

2. Sers R., Forrester S., Moss E., Ward S., Ma J., Zecca M. Validity of the Perception Neuron Inertial Motion Capture System for Upper Body Motion Analysis. *Measurement*. 2020;149:107024. http://dx.doi.org/10.1016/j.measurement.2019.107024

3. Bauer P., Lienhart W., Jost S. Accuracy Investigation of the Pose Determination of a VR System. *Sensors*. 2021;21(5):1622. http://dx.doi.org/10.3390/s21051622

4. Irshad M.T., Nisar M.A., Gouverneur P., Rapp M., Grzegorzek M. AI Approaches towards Prechtl's Assessment of General Movements: A Systematic Literature Review. *Sensors*. 2020;20(18):5321. http://dx.doi.org/10.3390/s20185321

5. Merriaux P., Dupuis Y., Boutteau R., Vasseur P., Savatier X. A Study of Vicon System Positioning Performance. *Sensors*. 2017;17(7):1591. https://doi.org/10.3390/s17071591

6. Nakano N., Sakura T., Ueda K., Omura L., Kimura A., Iino Y., et al. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose with Multiple Video Cameras. *Frontiers in Sports and Active Living*. 2020;2:50. https://doi.org/10.3389/fspor.2020.00050

7. Coronado E., Fukuda K., Ramirez-Alpizar I.G., Yamanobe N., Venture G., Harada K. Assembly Action Understanding from Fine-Grained Hand Motions, a Multi-camera and Deep Learning Approach. In: *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* New York, NY: IEEE; 2021. P. 2628–2634. http://dx.doi.org/10.1109/IROS51168.2021.9636715

8. Tausif Diwan, Anirudh G., Tembhurne J.V. Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications. *Multimedia Tools and Applications*. 2023;82(6):9243–9275. https://doi.org/10.1007/s11042-022-13644-y

9. Wei Liu, Anguelov D., Erhan D., Szegedy C., Reed S., Cheng-Yang Fu., et al. SSD: Single Shot MultiBox Detector. In book: Leibe B., Matas J., Sebe N., Welling M. (eds). *Computer Vision – ECCV 2016*. Cham: Springer. 2016;9905:21–37. https://doi.org/10.1007/978-3-319-46448-0_2

10. Bharati P., Pramanik A. Deep Learning Techniques—R-CNN to Mask R-CNN: A Survey. In book: Das A., Nayak J., Naik B., Pati S., Pelusi D. (eds). *Computational Intelligence in Pattern Recognition*. New York, NY: Springer. 2020;999:657–668. http://dx.doi.org/10.1007/978-981-13-9042-5_56

11. Bajpai R., Joshi D. MoveNet: A Deep Neural Network for Joint Profile Prediction across Variable Walking Speeds and Slopes. *IEEE Transactions on Instrumentation and Measurement.* 2021;70:1–11. http://dx.doi.org/10.1109/TIM.2021.3073720

12. Ghanbari S., Ashtyani Z.P., Masouleh M.T. User Identification Based on Hand Geometrical Biometrics Using Media-Pipe. In: *Proc. 30th International Conference on Electrical Engineering (ICEE)*. New York, NY: IEEE; 2022. P. 373–378. http://dx.doi.org/10.1109/ICEE55646.2022.9827056

13. Weijian Mai, Fengjie Wu, Ziqian Guo, Yuhan Xiang, Gensheng Liu, Xiaobin Chen. A Fall Detection Alert System Based on Lightweight Openpose and Spatial-Temporal Graph Convolution Network. *Journal of Physics: Conference Series*. 2021;2035:012036. http://dx.doi.org/10.1088/1742-6596/2035/1/012036

14. Szeliski R. Recognition. In book: *Computer Vision: Algorithms and Applications*. London: Springer; 2011. P. 575–640. https://doi.org/10.1007/978-1-84882-935-0_14

15. Kahl F., Hartley R. Multiple-View Geometry Under the L∞-Norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008;30(9):1603–1617. http://dx.doi.org/10.1109/TPAMI.2007.70824

16. Luhmann T., Fraser C., Maas H.-G. Sensor Modelling and Camera Calibration for Close-Range Photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016;115:37–46. https://doi.org/10.1016/j.isprsjprs.2015.10.006

17. Кудинов И.А., Павлов О.В., Холопов И.С. Реализация алгоритма определения пространственных координат и угловой ориентации объекта по реперным точкам, использующего информацию от одной камеры. *Компьютерная оптика.* 2015;39(3):413–419. https://doi.org/10.18287/0134-2452-2015-39-3-413-419

18. Jen-Li Chung, Lee-Yeng Ong, Meng-Chew Leow. Comparative Analysis of Skeleton-Based Human Pose Estimation. *Future Internet*. 2022;14(12):380. https://doi.org/10.3390/fi14120380

19. Jia Chen, Dongli Wu, Peng Song, Fuqin Deng, Ying He Y., Shiyan Pang. Multi-View Triangulation: Systematic Comparison and an Improved Method. *IEEE Access*. 2020;8:21017–21027. http://dx.doi.org/10.1109/ACCESS.2020.2969082

20. Obukhov A., Dedov D., Volkov A., Teselkin D. Modeling of Nonlinear Dynamic Processes of Human Movement in Virtual Reality Based on Digital Shadows. *Computation*. 2023;11(5):85. https://doi.org/10.3390/computation11050085

21. Abella J., Demircan E. A Multi-Body Simulation Framework for Live Motion Tracking and Analysis within the Unity Environment. In: *Proc. 16th International Conference on Ubiquitous Robots (UR)*. New York, NY: IEEE; 2019. P. 654–659. http://dx.doi.org/10.1109/URAI.2019.8768659

22. Zago M., Luzzago M., Marangoni T., De Cecco M., Tarabini M., Galli M. 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. *Frontiers in Bioengineering and Biotechnology*. 2020;8:181. https://doi.org/10.3389/fbioe.2020.00181

23. Latorre J., Llorens R., Colomer C., Alcañiz M. Reliability and Comparison of Kinect-Based Methods for Estimating Spatiotemporal Gait Parameters of Healthy and Post-Stroke Individuals. *Journal of Biomechanics*. 2018;72:268–73. https://doi.org/10.1016/j.jbiomech.2018.03.008

24. Обухов А.Д., Волков А.А., Вехтева Н.А., Патутин К.И., Назарова А.О., Дедов Д.Л. Метод формирования цифровой тени процесса перемещения человека на основе объединения систем захвата движений. *Информатика и автоматизация.* 2023;22(1):168–189. https://doi.org/10.15622/ia.22.1.7

References

1. Lind CM, Abtahi F, Forsman M. Wearable Motion Capture Devices for the Prevention of Work-Related Musculoskeletal Disorders in Ergonomics – An Overview of Current Applications, Challenges, and Future Opportunities. *Sensors*. 2023;23(9):4259. https://doi.org/10.3390/s23094259

2. Sers R, Forrester S, Moss E, Ward S, Ma J, Zecca M. Validity of the Perception Neuron Inertial Motion Capture System for Upper Body Motion Analysis. *Measurement*. 2020;149:107024. http://dx.doi.org/10.1016/j.measurement.2019.107024

3. Bauer P, Lienhart W, Jost S. Accuracy Investigation of the Pose Determination of a VR System. *Sensors*. 2021;21(5):1622. http://dx.doi.org/10.3390/s21051622

4. Irshad MT, Nisar MA, Gouverneur P, Rapp M, Grzegorzek M. AI Approaches towards Prechtl's Assessment of General Movements: A Systematic Literature Review. *Sensors*. 2020;20(18):5321. http://dx.doi.org/10.3390/s20185321

5. Merriaux P, Dupuis Y, Boutteau R, Vasseur P, Savatier X. A Study of Vicon System Positioning Performance. *Sensors*. 2017;17(7):1591. https://doi.org/10.3390/s17071591

6. Nakano N, Sakura T, Ueda K, Omura L, Kimura A, Iino Y, et al. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose with Multiple Video Cameras. *Frontiers in Sports and Active Living*. 2020;2:50. https://doi.org/10.3389/fspor.2020.00050

7. Coronado E, Fukuda K, Ramirez-Alpizar IG, Yamanobe N, Venture G, Harada K. Assembly Action Understanding from Fine-Grained Hand Motions, a Multi-camera and Deep Learning Approach. In: *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* New York, NY: IEEE; 2021. P. 2628–2634. http://dx.doi.org/10.1109/IROS51168.2021.9636715

8. Tausif Diwan, Anirudh G, Tembhurne JV. Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications. *Multimedia Tools and Applications*. 2023;82(6):9243–9275. https://doi.org/10.1007/s11042-022-13644-y

9. Wei Liu, Anguelov D, Erhan D, Szegedy C, Reed S, Cheng-Yang Fu, et al. SSD: Single Shot MultiBox Detector. In book: Leibe B, Matas J, Sebe N, Welling M. (eds). *Computer Vision – ECCV 2016*. Cham: Springer. 2016;9905:21–37. https://doi.org/10.1007/978-3-319-46448-0_2

10. Bharati P, Pramanik A. Deep Learning Techniques—R-CNN to Mask R-CNN: A Survey. In book: Das A., Nayak J., Naik B., Pati S., Pelusi D. (eds). *Computational Intelligence in Pattern Recognition*. New York, NY: Springer. 2020;999:657–668. http://dx.doi.org/10.1007/978-981-13-9042-5_56

11. Bajpai R, Joshi D. MoveNet: A Deep Neural Network for Joint Profile Prediction across Variable Walking Speeds and Slopes. *IEEE Transactions on Instrumentation and Measurement.* 2021;70:1–11. http://dx.doi.org/10.1109/TIM.2021.3073720

12. Ghanbari S, Ashtyani ZP, Masouleh MT. User Identification Based on Hand Geometrical Biometrics Using Media-Pipe. In: *Proc. 30th International Conference on Electrical Engineering (ICEE)*. New York, NY: IEEE; 2022. P. 373–378. http://dx.doi.org/10.1109/ICEE55646.2022.9827056

13. Weijian Mai, Fengjie Wu, Ziqian Guo, Yuhan Xiang, Gensheng Liu, Xiaobin Chen. A Fall Detection Alert System Based on Lightweight Openpose and Spatial-Temporal Graph Convolution Network. *Journal of Physics: Conference Series*. 2021;2035:012036. http://dx.doi.org/10.1088/1742-6596/2035/1/012036

14. Szeliski R. Recognition. In book: *Computer Vision: Algorithms and Applications*. London: Springer; 2011. P. 575–640. https://doi.org/10.1007/978-1-84882-935-0_14

15. Kahl F, Hartley R. Multiple-View Geometry Under the L∞-Norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008;30(9):1603–1617. http://dx.doi.org/10.1109/TPAMI.2007.70824

16. Luhmann T, Fraser C, Maas H-G. Sensor Modelling and Camera Calibration for Close-Range Photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016;115:37–46. https://doi.org/10.1016/j.isprsjprs.2015.10.006

17. Kudinov IA, Pavlov OV, Holopov IS. Implementation of an Algorithm for Determining the Spatial Coordinates and the Angular Orientation of an Object Based on Reference Marks Using Information from a Single Camera. *Computer Optics*. 2015;39(3):413–419. https://doi.org/10.18287/0134-2452-2015-39-3-413-419

18. Jen-Li Chung, Lee-Yeng Ong, Meng-Chew Leow. Comparative Analysis of Skeleton-Based Human Pose Estimation. *Future Internet*. 2022;14(12):380. https://doi.org/10.3390/fi14120380

19. Jia Chen, Dongli Wu, Peng Song, Fuqin Deng, Ying He Y, Shiyan Pang. Multi-View Triangulation: Systematic Comparison and an Improved Method. *IEEE Access*. 2020;8:21017–21027. http://dx.doi.org/10.1109/ACCESS.2020.2969082

20. Obukhov A, Dedov D, Volkov A, Teselkin D. Modeling of Nonlinear Dynamic Processes of Human Movement in Virtual Reality Based on Digital Shadows. *Computation*. 2023;11(5):85. https://doi.org/10.3390/computation11050085

21. Abella J, Demircan E. A Multi-Body Simulation Framework for Live Motion Tracking and Analysis within the Unity Environment. In: *Proc. 16th International Conference on Ubiquitous Robots (UR)*. New York, NY: IEEE; 2019. P. 654–659. http://dx.doi.org/10.1109/URAI.2019.8768659

22. Zago M, Luzzago M, Marangoni T, De Cecco M, Tarabini M, Galli M. 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. *Frontiers in Bioengineering and Biotechnology*. 2020;8:181. https://doi.org/10.3389/fbioe.2020.00181

23. Latorre J, Llorens R, Colomer C, Alcañiz M. Reliability and Comparison of Kinect-Based Methods for Estimating Spatiotemporal Gait Parameters of Healthy and Post-Stroke Individuals. *Journal of Biomechanics*. 2018;72:268–273. https://doi.org/10.1016/j.jbiomech.2018.03.008

24. Obuhov AD, Volkov AA, Vekhteva NA, Patutin KI, Nazarova AO, Dedov DL. The Method of Forming a Digital Shadow of the Human Movement Process Based on the Combination of Motion Capture Systems. *Informatics and Automation*. 2023;22(1):168–189. https://doi.org/10.15622/ia.22.1.7

Поступила в редакцию 18.05.2023
Поступила после рецензирования 13.06.2023
Принята к публикации 19.06.2023

Об авторах:

Артём Дмитриевич Обухов, доктор технических наук, доцент кафедры системы автоматизированной поддержки принятия решений Тамбовского государственного технического университета, (392036, РФ, г. Тамбов, ул. Ленинградская, 1), ResearcherID, ScopusID, AuthorID, ORCID, obuhov.art@gmail.com

Денис Леонидович Дедов, кандидат технических наук, старший научный сотрудник Тамбовского государственного технического университета, (392000, РФ, г. Тамбов, ул. Советская, 116), ScopusID, AuthorID, ORCID, hammer68@mail.ru

Екатерина Олеговна Суркова, студентка 4 курса кафедры системы автоматизированной поддержки принятия решений Тамбовского государственного технического университета, (392036, РФ, г. Тамбов, ул. Ленинградская, 1), ScopusID, AuthorID, ORCID, esur2506@yandex.ru

Ирина Львовна Коробова, кандидат технических наук, заведующая кафедрой системы автоматизированной поддержки принятия решений Тамбовского государственного технического университета, (392036, РФ, г. Тамбов, ул. Ленинградская, 1), ScopusID, AuthorID, ORCID, ira.sapr.tstu@mail.ru

Заявленный вклад соавторов:

А.Д. Обухов — формирование основной концепции, научное руководство, формулирование цели и задач исследования, подготовка текста, формирование выводов.

Д.Л. Дедов — организация экспериментальных исследований, подготовка текста.

Е.О. Суркова — проведение экспериментальных исследований, разработка программного обеспечения.

И.Л. Коробова — анализ результатов исследования, доработка текста.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Все авторы прочитали и одобрили окончательный вариант рукописи

Received 18.05.2023
Revised 13.06.2023
Accepted 19.06.2023

About the Authors:

Artem D. Obukhov, Dr.Sci. (Eng.), Associate Professor of the Department of Automated Systems for Decision-Making Support, Tambov State Technical University (1, Leningradskaya St., Tambov, 392036, RF), ResearcherID, ScopusID, AuthorID, ORCID, obuhov.art@gmail.com

Denis L. Dedov, Cand.Sci. (Eng.), Senior Researcher, Tambov State Technical University (116, Sovetskaya St., Tambov, 392000, RF), ScopusID, AuthorID, ORCID, hammer68@mail.ru

Ekaterina O. Surkova, 4th year student of the Department of Automated Systems for Decision-Making Support, Tambov State Technical University (1, Leningradskaya St., Tambov, 392036, RF), ScopusID, AuthorID, ORCID, esur2506@yandex.ru

Irina L. Korobova, Cand.Sci. (Eng.), Head of the Department of Automated Systems for Decision-Making Support, Tambov State Technical University (1, Leningradskaya St., Tambov, 392036, RF), ScopusID, AuthorID, ORCID, ira.sapr.tstu@mail.ru

Claimed contributorship:
AD Obukhov: basic concept formulation, academic advising, formulation of the research objective and tasks, text preparation, formulation of conclusions.
DL Dedov: organization of experimental research, text preparation.
EO Surkova: experimental research, software development.
IL Korobova: analysis of research results, revision of the text.

Conflict of interest statement: the authors do not have any conflict of interest.

All authors have read and approved the final version of the manuscript.